Chapter 6
Data Management and Archiving

The Data Management Plan (DMP; Mortenson 2006) and Vital Signs Monitoring Plan, although separate documents, were prepared simultaneously and are conceptually linked. This chapter presents an overview of the DMP and outlines the steps that SWAN will follow in managing and disseminating data that are acquired from long-term ecological monitoring. Information is the common currency among the activities and staff involved in the stewardship of natural resources for the NPS. This chapter summarizes the SWAN data management strategy, which is more fully presented in the SWAN DMP. The DMP is a guide for current and future project leaders and Network staff to ensure the continuity and documentation of data management methods and procedures over time. The DMP, in turn, refers to other guidance documents and standard operating procedures that convey the specific standards and steps for achieving the data management goals.

The DMP focuses on the processes used to:

1. Acquire, store, manage, and archive data
2. Ensure data quality
3. Document, analyze, summarize, and disseminate data
4. Ensure the long-term access to and utility of data.

6.1 Data Management Goals

The goal of the NPS Inventory & Monitoring (I&M) Program is to provide scientifically and statistically sound data to support management decisions for the protection of park resources. The goal of data management is to ensure the quality, interpretability, security, longevity, and availability of our natural resource data. The goal of the DMP is to outline the procedures and work practices that support effective data management.

The DMP objectives are to ensure that:

- **Data managed by the Network are of high quality**, including designing standardized data entry, importation, and handling procedures that effectively screen for inappropriate data and minimize transcription and translation errors;
- **Network data can be easily interpreted**, by considering the users' needs as the primary factor driving the design of summary reports and analyses; establishing rigorous data documentation standards; integrating common data tables and fields in NPS or regional standards; and making summary information available in formats tailored to the variety of audiences interested in I&M program results;
- **Data are secure for the long term**, including instituting standard procedures for versioning, data storage, and archiving; and natural history archiving, curation, and records management are provided to NPS curators;
- **Network data are readily available**, by implementing standard procedures for distributing data, while protecting sensitive data and by designing a standardized filing system for organizing I&M information.

6.2 Type of Information Managed by SWAN

The term "data" is frequently used in a way that also encompasses other products generated alongside the tabular and spatial data that are the primary targets of our data management efforts. These products fall into general categories listed in Table 6-1.

Table 6-1 *General categories of data products.*

Category	Description	Examples
1) Data		
• ***Raw Data***	Data obtained from the environment and that has not been subjected to any quality assurance or control beyond those applied during field work.	• field data sheets • specimens • remotely sensed data • data gathered electronically on field computers • GPS rover files • photographic imagery
• ***Validated and Verified Data***	Data that have been verified according to the standard operating procedure under which the data were gathered (typically the protocol for a given monitoring component) and are deemed ready for reporting and/or analysis.	• Relational databases • Tabular data files • Laboratory results • GIS layers • Maps • Processed or analyzed remote sensing data
• ***Analyzed Data***	Data that have been subjected to analytical routines after field collection and verification. This includes statistical operations conducted on the data for the purposes of arriving at a measure of the given ecological parameter or a compilation of analyzed data from different sources or time periods to derive new information.	• Summarized reports, data and maps from statistical or query operations • GIS/maps derived or repeated from remote sensing data. • Multimedia products, such as videos or slideshows
2) Documentation	Documentation provides the information required to understand the context of the data.	• Data collection protocols • Data processing/analysis protocols • Record of protocol changes • Data dictionary • FGDC metadata • Database design documents • QA/QC reports • Catalogs
3) Reports	Reports provide a means of presenting and publishing the methods and the results of analysis in the context of which it was intended.	• Annual progress reports • Final reports • Trend analysis reports • Publications • Final data posted on websites
4) Administrative Records	Administrative records supplement the context of a project and should be considered part of the projects deliverables.	• Contracts and agreements • Study and work plans • Research permit • Critical administrative correspondence

6.3 Priorities of Natural Resource Data

The priorities for Network data management efforts are:

- Produce and curate high-quality, well-documented data originating with the I&M Program
- Assist with data management for current projects, legacy data, and data originating outside the I&M Program that complement program objectives
- Help ensure good data management practices for park-based natural resource projects that are just beginning to be developed and implemented.

6.4 Data Stewardship Roles and Responsibilities

Every individual involved in the I&M Program is required to understand and perform data stewardship responsibilities in the production, analysis, management, and end use of the data as described in the DMP and the specific monitoring protocols. Specific roles and responsibilities for vital signs monitoring are written in each monitoring protocol. Senior staff (described in Chapter 8) share the responsibility in ensuring that data management procedures are followed (see Figure 6-1).

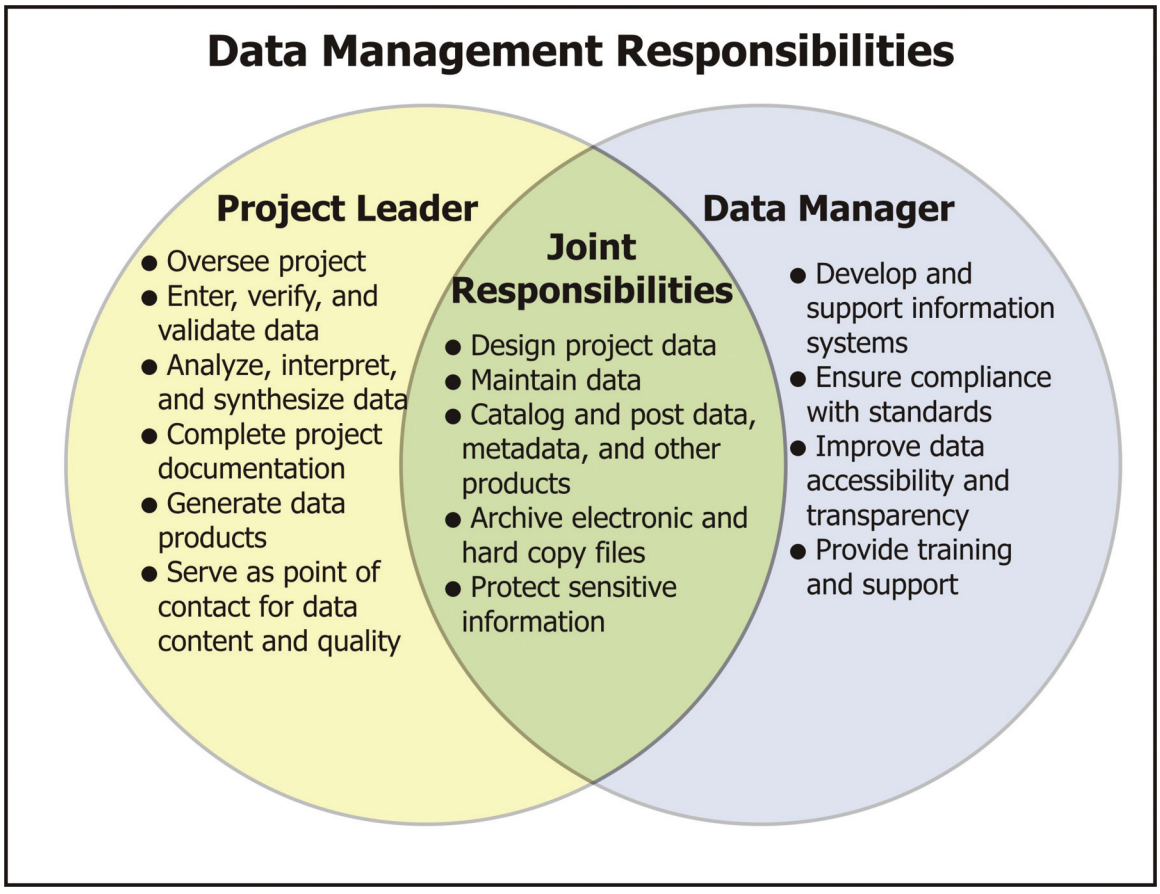

Figure 6-1 *Core project data stewardship duties of project leaders and data managers.*

6.5 Information Work Flow

Understanding the life cycle of data throughout a project will help to manage the staffing resources necessary to complete and support quality data. For data management to be effective, it must occur throughout the project life cycle.

A project is divided here into the following stages (see Figure 6-2):

1. Project initiation
2. Planning and approval
3. Design and testing
4. Implementation
5. Product delivery
6. Product integration
7. Closure and evaluation

SWAN uses a project tracking database to document and support the progress of information collected for vital signs monitoring. Most notably, this database tracks the status, changes, archiving and distribution of deliverables.

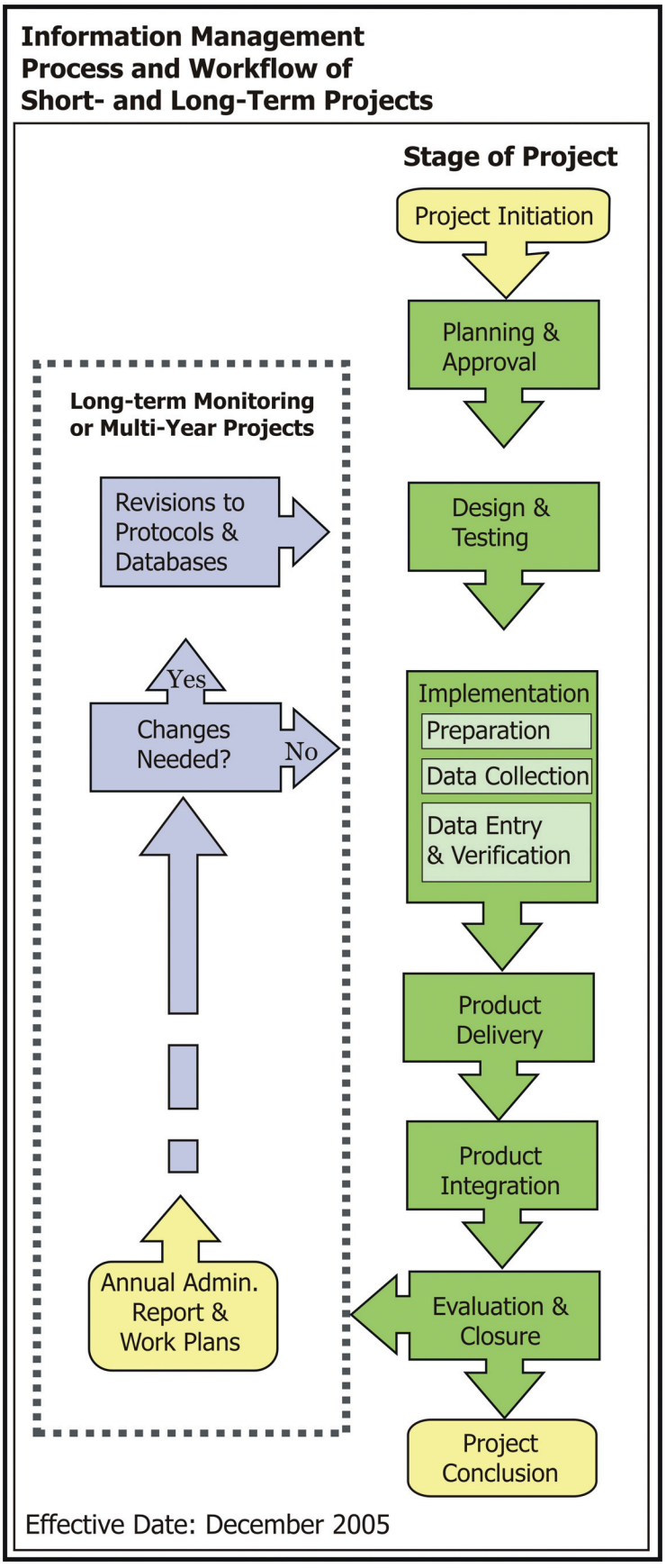

Figure 6-2 *Workflow overview.*

6.6 Information Technology (IT) Infrastructure for Data Management

Infrastructure refers to the network of computers and servers that our information systems are built upon. SWAN relies heavily on the national, regional, and park IT personnel and resources to maintain its computer infrastructure. This includes, but is not limited to: computers, servers, and other related hardware; software installation and support; e-mail administration; security updates; virus protection; telecommunications; computer networking; and backups of servers.

The infrastructure needs to support these required functions:

- Provide a central repository for master data sets
- Provide controlled subsets of data for local computing
- Provide a means for uploading and downloading data for both the NPS and the public
- Support desktop and internet applications
- Provide security, stability, and backups.

SWAN will also utilize infrastructure and information systems from partners who share monitoring objectives or provide source information. Anticipated partners are listed in Chapter 8 (Section 8.4).

6.7 Database Design Strategies

The project leader and the data manager will work together to develop conceptual data models to:

- Understand conceptually the data life cycle flow of the data collection process, e.g., where is the starting point of the data collection (for example, a visit to a site) and what happens next.
- Determine the data relationships as the implementation progresses, e.g., one site visited many times with many collections.
- Determine how the information will be presented.

Understanding the relationships between the data components collected is key to the success of a database and its utility. If the relationships are misunderstood, the database may become tedious in data entry and cumbersome at data output.

The SWAN DMP specifies the standards by which data will be handled. Data management elements or principles common to more than one vital sign will be managed in a conventional manner to allow for greater comparison of data across the Network, as well as to ensure further general data integrity.

6.8 Acquiring and Processing Data

The types of data handled by the I&M Program fall into three general categories:

- Program data—produced by projects that are either initiated (funded) by the I&M Program or involve the I&M Program in another manner (e.g., natural resource inventories and vital signs monitoring projects)
- Nonprogram legacy/existing data—produced by NPS entities without the involvement of the I&M Program (e.g., park or regional projects)
- Nonprogram external data—produced by agencies or institutions other than the NPS (e.g., weather and water quality data).

Most data acquired by the Network will be collected as field data (inventories and monitoring studies) or discovered through data mining initiatives (legacy/existing data). Methods of field data collection, such as paper filed data forms, field computers, automated data loggers, and GPS units will be specified in individual monitoring protocols and study plans. Field crew members will closely follow the established standard operating procedures (SOPs) in the project protocol. Data acquired by non-program sources, such as data downloaded from other agencies, will also be specified in individual monitoring protocols.

6.9 Ensuring Data Quality

The effort to detect trends and patterns in ecosystem processes requires data of documented quality that minimize error and bias. High quality data and information are vital to the credibility and success of the I&M Program, and everyone plays a part in ensuring that products conform to data quality standards.

Although many quality assurance/quality control (QA/QC) procedures depend upon the individual vital signs being monitored, some general concepts apply to all. Specific procedures to ensure data quality must be included in the protocols for each vital sign. Examples of QA/QC practices include:

- Field crew training
- Standardized field data sheets with descriptive data dictionaries
- Use of handheld computers and data loggers
- Equipment maintenance and calibration
- Procedures for handling data in the field
- Database features to minimize transcription errors, including imports from data loggers, range limit, pick lists, etc.
- Verification and validation, including automated error-checking database routines

QA methods should be in place at the inception of any project and continue through all project stages to final archiving of the data set. It is critical that each member of the team work to ensure data quality.

The final step in project QA is the preparation of summary documentation that assesses the overall data quality. A statement of data quality will be composed by the project leader and incorporated into formal metadata. Metadata for each data set will also provide information on the specific QA procedures applied and the results of the review.

6.10 Data Documentation

Documenting data sets, data sources, and methodology by which the data were acquired establishes the basis for interpreting and appropriately using data. At a minimum, all data managed by the Network will require the following elements of documentation:

- Project documentation
- Formal metadata compliant with Federal Geographic Data Committee (FGDC) standards
- Data dictionaries and entity relationship diagrams for all tabular databases.

Data documentation will be available and searchable in conjunction with related data and reports via the SWAN Web site as well as with the NPS Natural Resource and GIS Programs metadata and data store (NR-GIS Metadata and Data Store).

6.11 Summarizing and Analyzing Data

Providing meaningful results from data summary and analysis is a cornerstone of the I&M Program and characterizes the Network's data management mission to provide useful information for managers and scientists. Each monitoring protocol establishes requirements for on-demand and scheduled data analysis and reporting. Based on these requirements, the associated databases for the protocols include functions to summarize and report directly from the database as well as output formats for import to other analysis software programs. In addition to tabular and charted summaries, the Network provides maps of natural resource data and geographic information system (GIS) analysis products to communicate spatial locations, relationships, and geospatial model results. See Chapter 7 for a more detailed description of the Network's analysis and reporting schedule and procedures.

6.12 Data Dissemination

The SWAN data dissemination strategy aims to ensure that:

- Data are easily discoverable and obtainable.
- Only data subjected to complete QC are released, unless necessary in response to a Freedom of Information Act (FOIA) request.
- Distributed data are accompanied by appropriate documentation.
- Sensitive data are identified and protected from unauthorized access and inappropriate use.

Access to SWAN data products will be facilitated via a variety of means that allow users to browse, search, and acquire Network data and supporting documents. These means include, but are not limited to:

- SWAN public Web site, under "Information Discovery" (http://www.nature.nps.gov/im/units/swan/)
- NR-GIS Metadata and Data Store. Distribution instructions for each data set will be provided in the respective metadata. The NR-GIS Metadata and Data Store is available at http://science.nature.nps.gov/nrdata/, and is also accessible by the National Spatial Data Infrastructure, Geospatial One Stop Web site (http://gos2.geodata.gov).
- Alaska Geographic Data Committee Web site
- Service-wide databases, such as NPSTORET, NPSpecies, and NatureBIB
- Regional, Network, or park data servers protected with read-only access
- External repositories such as the Alaska Resource Library and Information Service, U.S. Geological Survey, University of Alaska, Alaska Department of Fish and Game, Western Regional Climatic Center, Exxon Valdez Oil Spill Trustee Council, and many others
- FTP sites, CDs, or DVDs, as appropriate.

6.13 Ownership, FOIA, and Sensitive Data

SWAN products are considered property of the NPS. However, FOIA establishes access by any person to federal agency records that are not protected from disclosure by exemption or by special law enforcement record exclusions. The NPS is directed to protect information about the nature and location of sensitive park resources under one Executive Order and four resource confidentiality laws:

- Executive Order No. 13007: Indian Sacred Sites
- National Parks Omnibus Management Act (NPOMA; 16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

When any of these regulations are applicable, public access to data can be restricted. If disclosure could result in harm to natural resources, the records may be classified as "protected" or "sensitive" and information may be withheld regarding the following resources recognized as sensitive by the NPS:

- Endangered, threatened, rare, or commercially valuable National Park System resources
- Mineral or paleontological sites
- Objects of cultural patrimony
- Significant caves

The Network will comply with all FOIA restrictions regarding the release of data and information, as instructed in NPS Director's Order 66 and accompanying Reference Manuals 66A and 66B (currently in development). Managing natural resource information that is sensitive or protected requires the following steps:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of which data must not be released in a public forum
- Management and archiving of those records to avoid their unintentional release

Classification of sensitive data will be the responsibility of Network staff, park superintendents, and project leaders. Network staff will classify sensitive data on a case-by-case, project-by-project basis and will work closely with project leaders to ensure that potentially sensitive park resources are identified, that information about these resources is tracked throughout the project, and that potentially sensitive information is removed from documents and products that will be released outside the Network.

6.14 Data Maintenance, Storage, and Archiving

SWAN data maintenance, storage, and archiving procedures aim to ensure that data and related documents (digital and analog) are:

- Kept up to date with regards to content and format such that the data are easily accessed and their heritage and quality are easily learned.
- Physically secure against environmental hazards, catastrophe, and human malice.

Primary data maintenance will be performed on the central Alaska Regional Office server or Network server and will follow the regional office's backup procedures. Data and information content of SWAN files stored on this server will be kept current. Accompanying documentation files will reflect any updates. These information files will be properly cataloged and maintained on the SWAN Web site. Latest versions of primary data will be available in conventional formats reflecting common data usages in the resource management community.

Project data will be electronically archived as stand-alone products and will include:

- Project documentation
- Data in raw, verified, and analyzed conditions
- Respective metadata
- Supporting files, such as photographs, maps, etc.
- All associated reports

Final deliverables from project data will be integrated with ongoing libraries and databases.

6.15 Natural History Archiving, Curation, and Records Management

In most instances, administrative documents, natural history specimens, photographs, audio tapes and other materials are essential companions to the digital data. Direction for managing many of these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records, that is, necessary for fulfillment of the NPS mission, and must be permanently archived.

The SWAN DMP includes a project checklist to guide project leaders in complying with archival directives. Physical items considered project products, such as reports, maps, photographs, or notebooks, will be cataloged and filed in the Network's central office and accessioned through the NPS Rediscovery curatorial database. A copy of the accessioned material will be archived according to NPS Standards and follow the procedures outlined in the SWAN DMP. Physical specimens, such as plants, animals, or tree core samples, will be accessioned and housed at the appropriate and accepted archival institution.

6.16 Water Quality Data

Water quality data are managed according to guidelines from the NPS Water Resources Division. This includes using the NPSTORET desktop database application to help manage data entry, documentation, and transfer. The Network oversees the use of NPSTORET according to the Network's integrated and regulatory water quality monitoring protocols and ensures that the content is transferred at least annually to the NPS Water Resources Division for upload to the STORET database.